

# PERBANDINGAN ALGORITMA CLUSTERING K-MEANS DAN K-MEDOID UNTUK PERINGKASAN TEKS OTOMATIS ARTIKEL BERITA

## LATAR BELAKANG



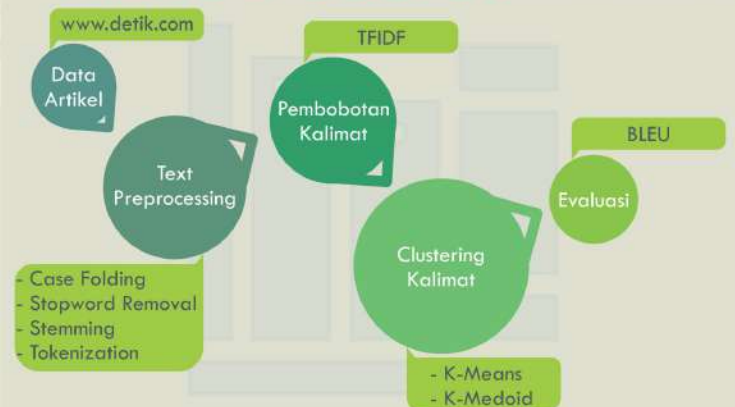
Kemajuan teknologi yang terus berkembang menghasilkan sejumlah informasi yang dapat diakses oleh kalangan masyarakat luas dengan mudah. World Wide Web menyediakan banyak sekali informasi di internet dimana setiap orang yang mencari akan mendapatkan informasi yang berbeda, sehingga akan mustahil bila setiap orang harus membaca keseluruhan informasi yang ada di internet. Terdapat beberapa contoh dari sekian informasi tersebut adalah data teks dokumen, yang berupa artikel berita, buku elektronik, dokumen penelitian, dan lain sebagainya yang membuat penyaringan informasi menjadi tidak efisien dan sulit dilakukan sehingga perlu adanya teknik otomatis untuk menyajikan informasi penting yang ringkas dan jelas. Dengan menggali informasi yang mendalam dapat membangkitkan informasi dari keseluruhan dokumen secara ringkas dan dapat memenuhi kebutuhan pembaca mengetahui dengan cepat isi dokumen tanpa harus membaca keseluruhan isi dokumen. Metode K-Means sendiri merupakan metode yang dapat mengenali pengelompokan secara alami dari suatu data yang tidak atau belum memiliki label tertentu serta pengelompokan dilakukan berdasarkan ukuran kemiripan antar data yang ada, dengan ukuran kemiripan yang tinggi antar data dalam satu kelompok dan kemiripan yang rendah dengan data yang berada pada kelompok lainnya. Serta metode K-Medoid sendiri adalah algoritma dengan kemampuan lebih kuat terhadap *noise* dan *outliner* apabila dibandingkan dengan K-Means karena lebih meminimalkan jumlah ketidaksamaan, bukannya meminimalkan jumlah kuadrat dari jarak *euclidean distance*.

## TUJUAN



Tujuan dari penelitian ini adalah membandingkan algoritma KMeans dan K-Medoid untuk mesin peringkasan teks otomatis artikel beritanya bersumber dari [www.detik.com](http://www.detik.com) dengan melihat perbedaan nilai akurasi yang dihasilkan oleh masing-masing algoritma.

## METODOLOGI



## HASIL PENELITIAN

Artikel yang mampu menghasilkan nilai tertinggi

No.	Compression Rate	K-Means	K-Medoid
1	10%	82,63%	81,69%
2	20%	70,62%	79,27%
3	30%	80,27%	80,27%

Hasil rata-rata peringkasan otomatis

No.	Compression Rate	K-Means	K-Medoid
1	10%	13,75%	12,74%
2	20%	30,59%	31,8%
3	30%	45,3%	44%



## KESIMPULAN

Hasil yang diperoleh dari membandingkan peringkasan teks otomatis dengan algoritma K-Means dan K-Medoid dapat disimpulkan bahwa algoritma K-Means lebih baik karena memiliki tingkat presisi yang tinggi pada nilai kompresi 30% dengan rata-rata nilai presisi 45,3%. Sedangkan algoritma K-Medoid mampu memperoleh nilai presisi yang tinggi pada nilai kompresi 30% dengan rata-rata nilai presisi 44%. Pada saat melakukan eksperimen diketahui bahwa panjang artikel dan nilai kompresi yang diambil sangat berpengaruh terhadap nilai presisi dan waktu komputasi yang dihasilkan. Semakin panjang artikel yang diringkaskan maka waktu komputasi akan bertambah lama meskipun dalam penelitian yang dilakukan masih memiliki waktu kurang dari 1 detik. Waktu komputasi akan terasa apabila dilakukan dengan data yang berjumlah cukup banyak.



## SARAN

Berdasarkan hasil dari penelitian yang telah dilakukan, peneliti dapat memberikan saran yang mungkin dapat membantu untuk penelitian ini dikembangkan kemudian hari, yaitu:

1. Dapat ditambahkan algoritma yang mampu membantu peringkasan teks otomatis multi dokumen
2. Dapat menggunakan algoritma klastering lainnya untuk peringkasan teks otomatis.
3. Dapat menggunakan metode *variance initialization* untuk mendapatkan *centroid* yang baik
4. Dapat menambah atau mengurangi nilai *compression rate* yang digunakan dalam uji coba.

Kata Kunci : Peringkasan Teks, K-Means, K-Medoid, BLEU



Rizqy Linarko



Setio Basuki, S.T., M.T



Nur Hayatin, S.T., M.Kom

